DOCUMENT RESUME

ED 205 591 TM 810 489

AUTHOR Sachse, Thomas P.
TITLE The Role of Performance Assessment in Tests of Problem Solving Ability.
INSTITUTION Northwest Regional Education Lab., Portland, Oreg. Clearinghouse for Applied Performance Testing.
SPONS AGENCY National Inst. of Education (ED), Washington, D.C.
PUB DATE Apr 81
GRANT 400-80-0105
NOTE 36p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Cognitive Tests; *Mastery Tests; Measurement Objectives; *Problem Solving; Scoring; *Test Format; Test Reliability; *Test Reviews; Test Validity

ABSTRACT

Consistent with its mission of synthesizing available information on performance assessment in a variety of skill areas, the Clearinghouse for Applied Performance Assessment has prepared this summary of the role of performance assessment in selected published tests of problem-solving skill. Performance assessment is defined in terms of assessment context, stimulus conditions, response mode, and scoring procedures, while problem solving is discussed as a psychological and educational construct. A historical review of strategies for the measurement of problem solving is followed by a synthesis and review of thirteen published tests of problem solving ability, with special emphasis on the role of performance assessment in these. Test profiles appear in the appendix, and are compared in terms of operational definition of problem solving, measurement strategies, response mode, and technical adequacy. In conclusion, three directions for future research are presented: (1) the determination of how general intelligence and problem solving can be used to understand real-life problem solving ability; (2) the development of simulated problem solving measures and; (3) the integration of problem solving activities with instructional programs, such as those dealing with the gifted, minimum competency students, and special education programs for the handicapped. (Author/AEF)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED205591

# THE ROLE OF PERFORMANCE ASSESSMENT IN TESTS OF PROBLEM-SOLVING ABILITY

Thomas P. Sachse
California Department of Education

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Prepared for

Clearinghouse for Applied Performance Testing
Northwest Regional Educational Laboratory
300 S.W. Sixth Avenue
Portland, OR 97204

April 1981

TM 810489

The material presented herein was developed by the Clearinghouse for Applied Performance Testing (CAPT) under grant #400-80-0105 with the National Institute of Education (NIE) of the Department of Education. The opinions expressed in this publication do not necessarily reflect the position of NIE and no official endorsement by the Institute should be inferred.

ABSTRACT

Consistent with its mission of synthesizing available information on performance assessment in a variety of skill areas, CAPT has prepared this summary of the role of performance assessment in selected published tests of problem-solving skill. Both performance assessment and problem solving are defined, and the relationship between the two is explored as it occurs in 13 published tests. Each test is profiled and the tests are compared in terms of operational definition of problem solving, measurement strategies, response mode, and technical adequacy.

## Introduction

The mission of the Clearinghouse for Applied Performance Testing (CAPT) includes a commitment to investigate the potential of performance assessment in the measurement of a variety of skills. Consistent with this mission, CAPT has prepared synthesis documents on the role of performance assessment in reading (Stiggins, 1980), writing (Spandel and Stiggins 1980), speaking and listening (CAPT, 1980), and performance assessment in medical education (Stiggins, 1979). This paper continues that synthesis effort by focusing on the role for performance assessment in selected published tests of problem-solving ability.

The paper focuses on four major topics. First, after a brief introduction to performance assessment, problem solving is discussed as a psychological and educational construct. Second, a historical review of strategies for the measurement of problem solving is presented. Third, some of the currently available published tests of problem-solving ability are reviewed and synthesized, with special emphasis on the role of performance assessment in those tests. And finally, conclusions are drawn regarding some potentially important directions for research and development of performance tests of problem-solving abilities.

## Performance Assessment

Performance assessment is best understood as a broad class of measurement options, rather than a very specific type of test. The class of options can be described in terms of four major attributes: assessment context, stimulus conditions, response mode, and scoring procedures.

The context of a performance assessment has two key dimensions. First, the purpose for using performance assessment is to measure examinees' ability to use available knowledge and skills (in this case, problem-solving skills) to achieve some relevant, real-world goal. Tests can be developed to measure (a) whether the examinee has mastered a given body of knowledge, or (b) is able to evaluate, synthesize and apply that knowledge to serve some meaningful purpose. Both are valuable uses of tests under certain circumstances; however, performance tests deal with the second purpose. To illustrate, science students might be tested on their knowledge of experimental method, or they might be tested on their ability to use that method to solve a problem such as determining the origin of some form of pollution in an ecology experiment. The latter is an example of performance assessment.

A second important aspect of performance assessment context is the setting in which assessment occurs. For example, tests may require the application of knowledge and skills in professional or occupational situations, personal life, or scholastic environments. Professional licensing examinations require the examinee to use information to solve problems confronted in professional situations. Functional literacy examinations often require the use of knowledge and skills to achieve personal life ends. And, science tests like the pollution experiment example given above require the use of knowledge in scholastic settings. Performance assessment exercises can be framed for all three settings.

Another major attribute of a performance test is the stimulus. The stimulus conditions of an assessment describe the specific problem for the examinee. Performance assessment can rely on either real or simulated stimulus conditions. For example, the stimulus might be provided by an actual work situation in which the examinee is required to demonstrate ability to carry out the correct job-related, problem-solving procedure (e.g., repairing a malfunctioning auto). Or, the stimulus might take the form of a simulated personal life problem (e.g., comparison shopping as part of consumer economics) in which stimulus circumstances are controlled (or set up) by the examiner to approximate the real world. A simulated condition can also take the form of a verbal (written) description of the problem circumstance.

The third important attribute of a performance test is the response mode. Here test developers have two choices. The examinee might be asked to select a response from among a finite list (e.g., multiple-choice test item), or generate an original response (e.g., writing sample). Since both objective test items and original responses can serve as a valid base for judging the examinee's ability to solve specified real-life problems, then logically both modes of response can be included in performance assessment options.

The final attribute of a performance test is the scoring mode. Here the alternatives vary along a continuum ranging from objective to subjective scoring. In objective scoring, the examinee's response is either correct or incorrect. The rater or scorer does not evaluate the degree of correctness. Objective test items are scored this way. Subjective scoring calls for a trained judge to observe the examinee's response and interpret its quality in terms of some internally held standards and criteria. The quality of the examinee's response may be evaluated as adequate or inadequate, or the judge may be called upon to discern varying degrees of adequacy.

Performance assessment can rely on objective or subjective scoring, since either can be used to evaluate the ability to use knowledge to achieve real-life goals. However, in CAPT's description, emphasis is given to the more subjective scoring of original responses.

To summarize, CAPT's description of performance assessment focuses on the application of knowledge and skills to achieve goals relevant to personal, professional or scholastic problem-solving situations in which real-life or simulated stimulus conditions elicit original responses that are scored subjectively by qualified judges. Now, let's examine how this form of assessment relates to the measurement of problem-solving ability.

## Definitions of Problem Solving

Interest in teaching and measuring problem-solving skill evolved from the work of Dewey (1910) and has been changing and growing ever since. Although there are numerous problem-solving definitions and models, three stand out as contributing most to the current understanding of this complex construct. These three include the definitions generated by Dewey (1910), Pribram (1960, 1968) and Guilford (1967, 1971).

Dewey's original work (1910) posited five steps in the problem-solving process:

1. A felt difficulty
2. Its location and definition
3. Suggestion of possible solution
4. Development by reasoning of the bearings of the suggestion
5. Further observation and experiment leading to its acceptance or rejection, that is, the conclusion of belief or disbelief

Dewey suggested that an educated mind was one that could estimate the utility of time spent on each process. Many others that followed also suggested componential or logically distinct stages of problem solving. Among them, Wallas (1912) proposed four phases: preparation, incubation, illumination and verification. Gagne (1964), in his early work on problem solving, listed five stages:

1. Reception of the stimulus situation
2. Concept formation or concept invention
3. Determining courses of action
4. Decision making
5. Verification

Feldhusen, Houtz and Ringenbach (1972) specified twelve activities in the problem-solving process:

1. Sensing that a problem exists
2. Defining the problem
3. Clarifying the goal
4. Asking questions
5. Guessing causes
6. Judging if more information is needed
7. Noticing relevant details
8. Using familiar objects in unfamiliar ways
9. Seeking implications
10. Solving single-solution problems
11. Solving multiple-solution problems and
12. Verifying solutions

An alternate paradigm for the conception of problem solving has been the information processing model. Pribram's work in this area (Miller, Galanter, and Pribram, 1960 and Pribram, 1971) centers on the cybernetic (qua, thermostatic) mechanism called TOTE (Test-Operate-Test-Exit). This information processing system is a serial representation of the problem solver as computer where a situation or condition is checked (test), manipulated (operate), rechecked (test), and satisfied (exit) when the original situation has been adequately dealt with. More elaborate information processing models of problem solving have developed based on extensive computer simulations. These include the work of Newell and Simon (Newell, Shaw, and Simon, 1958, 1962; Simon and Newell, 1971; Newell and Simon, 1972) that resulted in processes for generating a problem space and operating on the elements of that problem space.

Ausubel and Robinson (1969), Scandura (1977), and Gagne (1977), in his later work, have also proposed information processing models of problem solving. These models use varying conceptions of states, rules and processes that operate in uniform ways to deal with uncertainties and problems.

The work of Guilford in the area of problem solving might be considered a hybrid of the two categories described above. Guilford's "Structure of Intellect" (1967) presented an elaborate information processing model based on a constellation of human abilities organized into an holistic problem-solving approach. This model relies heavily on several input functions and the symbol systems in memory to produce and evaluate alternative problem solutions. Subsequent to the development of the Structure of Intellect, Guilford and Hoepfner (1971) used a series of abilities as well as a general problem-solving factor to represent the problem-solving process.

## Measurement of Problem Solving

Just as there are numerous conceptual models of problem solving, so too are there many measures of this complex construct. Reviews of these tests have been conducted by Ray (1955), Feldhusen, Houtz, Ringenbach and Lash (1971), Speedie, Treffingen, and Feldhusen (1973) and Shawn (1976). The Feldhausen et al. review grouped problem-solving tasks into four classes: (1) puzzle-insight, (2) process, (3) component, and (4) real-life problems. This categorization provides an excellent framework for discussing the historical and psychological development of problem-solving measures.

Puzzle-insight problems present novel situations that require juxtaposition or reorientation of stimuli to find the correct solution. These tasks often have binary scoring (right/wrong) with additional indices of problem-solving skill, such as time required for completion or number of hints required for solutions (Maier and Berke, 1966). Puzzle-insight problems include Maier's hatrack problems (1945), Luchins' water-jar problems (1942), Katona's matchstick problems (1940), and a variety of verbal analogy and anagram problems (Johnson, 1962, 1966).

In Maier's hatrack problems, the subject is given two sticks and a C-clamp where the hat rests. The solution to the problem requires that the problem solver clamp the two sticks together, wedging the ends against the ceiling and floor to create a standing hatrack. In Luchins' water-jar problems, the subject is given a number of jars (usually three) and is required to fill the largest and pour off enough water into the smaller vessels until a target remainder is left. For example, if 150 cc's of water are poured into the largest jar, and from that, 100 cc's go into the target remainder, the subject solves the problem by pouring 25 cc's into two other small containers. In the anagram problem, the subject receives stimulus letters printed on a 3" x 5" index card and is required to transpose letters until a target word is formed using only the letters in the stimulus word.

In all these problems, the key to finding a successful solution is for subjects to think flexibly about the problem and avoid being stymied by one apparent but inappropriate solution. Maier calls this mental block a "direction," Luchin refers to this as "Einstellung" (being blind to the solution), and Johnson labels this a mental "set" which must be broken before a solution is reached.

The two major limitations to these measures are that binary scoring reveals little about the problem-solving process, and that the measures have little relation to theoretical models of problem solving. The intermediate stage (referred to as direction, Einstellung or set) between the puzzle and the insight does not reveal much about the problem-solving act (Green, 1966). Investigations of problem solving by the Gestalt psychologists of the 40s and later focused primarily on the demographic (Bloom and Broder, 1950; Cunningham, 1966; Keisler, 1969) and situational (Burke, Maier and Hoffman, 1966; Burak and Moos, 1956; Corman, 1957) variables that helped subjects make the transition from puzzle to insight. Subsequent researchers (Cofer, 1957; Kendler & Kendler, 1962; Miles, 1968) posited theoretical models of problem solving and posed tasks that could empirically test such hypothetical processes.

Process problems were developed to overcome the limitations of the puzzle-insight tasks. As such, this class of problems illuminates the processes or steps employed by problem solvers by focusing on the patterns of problem solving rather than the solutions. Examples of process problems include switchlight problems (John, 1957; Tyler, 1958) and a host of simulation problems using the tab format (Glaser, Damrin, Gardener, 1954; Simon, 1970; and Rimoldi, 1960). The switchlight problem utilizes a problem-solving and information (PSI) apparatus; this is a board with nine lights in a circular array and a tenth light in the middle. Adjacent to each of the nine outside lights is a button which, when pressed, will activate (or deactivate) serially wired light around the board. The task for the subject is to illuminate the center (#10) light by pressing various combinations of buttons around the board.

The tab format is the forerunner of today's latent-image simulations. In the tab format, the subject responds yes/no to a variety of questions that give information to solve a problem. Each yes answer allows the subject to see the information under the associated tab. The subject is required to solve the problem using the most correct order and fewest number of tabs.

These problems provided an opportunity to create simulations in applied contexts, including medical diagnosis, military gaming and machine trouble-shooting. Here one begins to see the role of performance assessment in problem-solving tests. Another bonus of the process problems was the chance to test theoretical models of problem solving (Bruner, Goodnow and Austin, 1956; Davis, 1966).

Component problems represent the third class of problem-solving tests. Unlike the process problems which were used to generate the covert problem-solving steps, component problems were designed to measure discrete problem-solving skills. Numerous component problems were developed to validate various facets of the Structure of Intellect (Guilford, 1967). But the archetype of the component problems was a

series called the Torrance Test of Creative Thinking (1966). This battery was developed for school children (individually administered from kindergarten to fourth grade, group-administered after that). It has numerous subtests, each of which yields up to three types of scores: fluency, flexibility and originality. These subtests measure a variety of skills including Guilford's divergent and productive thinking.

The component problems succeeded in differentiating levels of problem solvers' abilities (Tate, Stanier, and Harootunian, 1959), but were unable to validate models such as Guilford's Structure of Intellect (1967) or the twelve problem-solving abilities espoused by Feldhusen, Houtz and Ringenbach (1972).

More recently, life skills tests of problem solving have been developed. This class of tests pose real-life problems for group or individual solution. Examples such as Crutchfield and Covington's (1966) The Man in the Pit and The Missing Jewel are beyond students' school experience. On the other hand, Treffinger's (1970) Fighting on the Playground, does tap students' experiential background. The life skills tests add the dimension of real-life relevance to the task, and sometimes the response, and are therefore examples of performance assessment. But they seldom relate to the problem-solving models requiring validation.

Having reviewed the problem-solving construct and the evolution of problem-solving measures, we can conclude that performance assessment has at least a potential role in problem solving. Now let's turn to the real world of published tests of problem-solving ability to determine the extent to which performance assessment is already playing a role.

## Published Problem-Solving Tests

Thirteen commercially available problem-solving tests were collected from publishers for review in this paper. From the Educational Testing Service lists entitled "Measures of Reasoning, Logical Thinking, and Problem-Solving Ability," those described as problem-solving tests were identified and copies were requested from the publisher. Of the approximately 20 test specimens requested, these 13 were received and reviewed:

Adult Performance Level Survey
Analysis of Relationships
Circus Test
Educational Goal Attainment Test and Assessment Planning
Learning Ability Profile
Means-Ends Problem Solving Procedure
Oregon Academic Ranking Test
Purdue Problem Solving Inventory
Ross Test of Higher Cognitive Processes
SRA Coping Skills
STS Educational Development Series
Stories About Real Life Problems
Watson-Glaser Critical Thinking Appraisal

Characteristics of each test are profiled in detail in the Appendix.

It is essential to remember that the reviewed tests include only those that are commercially available; research- and project-specific tests were not reviewed. The review covered six features of each instrument:

Conceptual Definition of Problem Solving
Specific Measurement Strategy
Role of Performance Assessment
Evidence of Reliability
Suggested Uses of the Test
Evidence of Validity for Intended Uses

In short, this review of problem-solving tests investigated the manner in which problem solving was assessed, the degree to which performance assessment played a part in the measurement process and the technical adequacy of each test. Each of these features is discussed, in turn, below.

Operational Definitions. In defining problem solving as a measurement construct, test developers use one of two styles: either they describe problem solving as one or more cognitive processes, or they list discrete problem-solving skills. Of the thirteen tests reviewed, eight are process oriented and five are skill oriented.

The skill oriented tests usually provide test specifications or blueprints that list each skill and the number of items measuring each skill. These tests include the life-skills measures. The Circus test, for example, lists five subparts on problem solving: these tests cover general knowledge, ability to detect incongruences, ability to define problems (by categorization and variation), ability to evaluate solutions, and ability to implement final decisions. Different item types and subtest scores are keyed to this problem-solving construct. Both group and individual diagnostic information for program planning are available.

Process-oriented tests tend to be more global in defining what is being assessed, and often measure intelligence or logical thinking in addition to problem-solving skills. In the Learning Ability Profile, for example, problem solving is a component of learning ability or general intelligence. According to the test manual, problem solving relates to "how one learns to learn, in that problems are placed against the background of prior experience." From this definition, test items measuring spatial and verbal abilities are employed to yield test scores for selection and placement decisions.

Another important dimension of the definitions for these tests is the degree to which they define problem solving. Eight tests give explicit definitions of the construct (whether process or skill oriented). These definitions usually include three to five components of problem solving, describe how those components are measured, and identify ways for using the data (for instructional diagnosis, program planning, etc.). The Means-Ends Problem Solving Procedure, for example, gives a concrete definition of problem solving that includes both cognitive and empathic understanding of the problem. However, five tests have implicit definitions for problem solving. In these tests, problem solving is

considered a subset of intelligence or learning ability. Most often those tests are used to place students in gifted programs or diagnose group needs for special instructional programs (of an unspecified nature). For the Oregon Academic Ranking Test, problem solving is operationally defined as academic or learning ability. This test measures "brightness" by assessing creativity and abstract thinking, and is suggested for use in placing gifted students.

Instrumentation. The test reviews focused in part on the fidelity of the stimuli and response modes to real life problem-solving skills. The stimuli employed vary in item type and content. The tests that have process-oriented definitions of problem solving often use abstract aptitude items that tap visual-spatial skills. The four tests with aptitude-type items require examinees to perform mental transformations of various relationships. Both visual and verbal objects are manipulated to describe categorizations, sequences and analogies; such is the case in the Analysis of Relationships battery. In these tests, multiple-choice items are used to predict academic and occupational success.

Two tests employ items that would be characterized as academic achievement items. These items measure examinees' prior experience and learning by asking knowledge and comprehension questions. These tests suggest that examinees who can answer these items correctly can make decisions based on information presented to them. Both tests are based on implicit, process-oriented definitions of problem solving.

A third group of tests uses life skills items to measure problem-solving ability. As mentioned earlier, these tests are the most recent development in the measurement of problem solving. The rationale for using real-life applications in testing problem-solving skills is twofold: first, examinee motivation is often higher in these types of problem-solving tests (Cronbach and Meele, 1955); and second, there is a growing interest among educators and the public in the application of basic skills to real-life problems.

The four tests with life skills items use everyday problems that require the synthesis, evaluation and application of information to solve a problem. For example, the Adult Performance Level Survey measures problem-solving ability in community resources, occupational knowledge, consumer economics, health and government and law. Life skills items on these tests are categorized by the content and/or skill being measured (e.g., consumer math, sequential synthesis). These tests claim that high scoring examinees demonstrate "coping skills" or functional competence. They clearly rely on higher fidelity stimuli than do the tests of abstract reasoning. Another group of four tests uses vignettes or simulations in the item stimuli. These tests present the examinee with a problem to be solved. The situations range from somewhat abstract (e.g., conceptualizing a problem) to very applied (e.g., solving an ecological problem), with the majority at the applied end of the continuum. These tests use novel presentation methods including slide-tape productions, film strips and complex diagrams. The use of simulations allows test developers to vary simultaneously the cognitive skill, content area and presentation mode. In the Means-Ends Problem-Solving procedure, ten vignettes are given to examinees. These vignettes portray various situations that require multiple steps for solution. Both the number of

steps and the quality of the solution figur- into the examinee's score. These are high fidelity tests.

Problem-solving tests that would be considered performance tests, according to CAPT's description of stimulus conditions include the following:

Adult Performance Level Survey
Educational Goal Attainment Test
Means-ends Problem-solving Procedures
Purdue Problem-solving Inventory
SRA Coping Skills
STS Educational Development Series
Stories About Real-life Problems
Watson-Glaser Critical Thinking Appraisal.

Regardless of the response mode (discussed next) these tests require the real-life application of problem-solving skills.

Response modes. The response modes for the 13 tests are far more limited than the stimuli. Eleven tests use multiple choice items, some with only three response options. It was surprising to find tests with life-skill or simulation items using multiple choice formats. However, since these tests are commercially published, the need for cost effectiveness may have required the use of machine scoring. A few of these 11 tests also use rating scales for attitudinal items. These tests assert that interest and emotional commitment are important components of problem solving.

The remaining two tests allowed examinees to generate novel responses to the applied stimuli. One open-response test allows for the oral presentation of the solution by the examinee (Means-ends Problem Solving Procedure), while the other open-response test allows for written responses (Oregon Academic Rank Test). In each case, a template or rubric is used to score the answers. The rubrics were not unlike those used in scoring writing samples. In these two cases, performance assessment is evident in the subjective rating methods used to evaluate original responses. The test developers felt the high fidelity of the performance component was worth the added cost of subjective scoring. These tests clearly have performance dimensions.

Technical adequacy. The technical information provided in the test manuals varies in completeness and technique. Reliability estimates are not provided for five of the tests. Two other tests give reliability estimates without revealing the technique used to obtain those estimates. The other six test manuals give reliability data based on the KR-20, split-half or odd-even reliability formulas. The total test reliabilities provided them range from .45 to .95, with the majority between .80 and .95. Those subtest reliabilities given range from .60 to .81 (reliable enough in some cases to make subtest interpretations for groups, but not individual students).

Evidence of validity is even more sparsely reported. There are no validity estimates for six of the 13 tests. Five test manuals list

concurrent validity estimates (with reading or intelligence tests). These estimates range from .45 (with performance ratings) to .65 (with reading test scores) to .70 (with intelligence tests). The Circus test gives a predictive validity coefficient of .83 (corrected for attenuation), as well as factor analytic evidence of construct validity.

By and large, however, evidence of technical adequacy for most of the tests--whether they include performance test components or not--is insufficient according to commonly accepted professional standards.

## Summary and Conclusions

This paper opened with a brief foray into problem solving as an educational and psychological construct. The various conceptions of problem solving described range from strings of mental steps, to information processing models, to human ability models.

The four classes of problem-solving instruments identified include some novel approaches to measurement. As a class, puzzle-insight tasks require subjects to solve a problem in which they have had little or no experience. Puzzle-insight problems help define situational and demographic characteristics affecting solution rates. The process problems, which include tab-format tasks, are forerunners of today's multifaceted simulation tests. These measures test complex skills without the realism of a hands-on performance test, and enhance our understanding of the steps used in problem solving. Component problems isolate specific problem-solving skills, and are often used to test creativity. Life skills tests measure problem solving in an applied context, renewing widespread interest in the technique and benefits of performance assessment.

Among the published tests reviewed for this analysis were tests that define problem solving as synonymous with intelligence and others that define it as a discrete list of mental skills. Some use objective items to measure academic knowledge while others present life skills vignettes and simulations. Among the published tests of problem solving reviewed, fully half fall under the rubric of performance assessment.

Problem solving is a concept with many meanings and many potential applications. Consequently, tests of problem-solving skill are designed for diverse purposes--including instructional management, selection, and program planning. Among the instructional management uses are diagnosis of student problem-solving strengths and weaknesses (via performance tests of steps in the problem-solving process), course placement, and educational/vocational guidance (using performance tests of life skills and functional problem-solving competence).

Among the selection uses are selection for admission into advanced and/or remedial programs, and certification of problem-solving competence, both of which can be conducted with tests that define problem solving globally or holistically. And among the program planning uses are formative program evaluation (with skill-oriented tests), and summative program evaluation or broad sample survey assessment, such as statewide assessment.

Because of scoring costs, performance tests are likely to be more useful in small-scale local classroom applications for instructional management and local program evaluation. For large-scale assessment, problem-solving tests that rely on objective test procedures are usually more cost effective.

This impressive array of uses shows that performance assessment can and does play a role in measuring problem-solving skills, and that problem-solving tests can be vital to many important educational decisions.

This is not to say that further research and development are not needed. Much remains to be learned about assessment of problem-solving skills. Three research and development priorities are discussed in the following paragraphs.

First, research should be conducted on how mental ability constructs like general intelligence and problem solving, can be used to understand real life problem-solving ability. It is important to find out whether aptitudes that are commonly assessed have any direct bearing on productive solutions to the problems encountered in daily life. For example, it may be that the problem-solving process encountered in one context is not the same as that required in another. Generalized problem-solving models may not relate to real life. But if context-specific models can be identified and validated, the skills (or steps) required to solve certain types of problems could be linked to learning programs which, in turn, would foster the growth of problem-solving skills among students. This research could have occupational and recreational applications (e.g., creating programs specifically for electronic technicians, horticulturists) as well as academic application.

Second, the development of simulated problem-solving measures should be continued and expanded. The tab format (of the process tests) has already been reborn in the simulation work of McGuire et al. (1976) for use in assessing skills of allied health professionals. McGuire's simulations employ latent-image techniques where a special pen brings out hidden answers to questions asked of medical assistants. This technique provides an almost limitless variety of tests customized by the examinee. Using this technique, problem-solving tests can use the branching options made popular in programmed and mastery instructional programs. These types of tests could also use extensive computer simulations to test problem-solving skills in many academic, occupational and personal settings.

Third, previous technical and measurement work in problem solving should be integrated with current instructional programs. At least three instructional areas could benefit from work in problem solving. Some tests reviewed in this paper could be very useful in gauging the progress of gifted students. Because of ceiling effects, typical standardized achievement tests tell little about such students' performance or potential. The tests reviewed here, on the other hand, could be used to estimate gifted students' problem-solving power (the time in which they can break the Einstellung).

At the other end of the instructional scale, minimum competency testing programs are being implemented around the nation to guarantee that low ability students meet a minimal level of proficiency before promotion or graduation. Public support of this testing movement reflects concern that all students get an adequate education so they can cope effectively in life outside school. The life skills problem-solving tests have real potential for tapping students' functional competence.

Special education programs are designed to help handicapped students live productive, fulfilling lives. Techniques used in developing problem-solving tests could be useful in creating simulations (as discussed above) or alternative modes of assessment for determining the effectiveness of such programs. When traditional paper and pencil tests cannot be used (e.g., by blind students), alternatives can be developed that allow for the assessment of important student abilities.

As educational research and development continues, special attention should be given to the increased test validity made possible through the use of performance assessment. This increase is most dramatically illustrated by life skills tests that incorporate performance testing techniques into the assessment of problem-solving skills and life skills.

REFERENCES


Ausubel, D.P., and Robinson, F.G. School learning: An introduction to educational psychology. New York, NY: Holt, Rinehart and Winston, 1969.

Bloom, B.S., and Broder, L.J. Problem-solving processes of college students: An exploratory investigation. Supplementary Education Monographs, Chicago, IL: University of Chicago, Press, 1950.

Bruner, J.S., Goodnow, Jacqueline, J., and Austin, G.A. A study of thinking. New York, NY: Wiley, 1956.

Burak, B., and Moos, J. Knowing the principle basic to solution of a problem. Journal of Educational Research, 1956, 50, 203-208.

Burke, R.J., Maier, N.R., and Hoffman, R. Functions of hints in individual problem-solving. American Journal of Psychology, 1966, Vol. 79, 389-399.

Clearinghouse for Applied Performance Testing. Assessing Speaking and Listening Skills, CAPTRENDS, Portland, OR: Northwest Regional Educational Laboratory, 6 (2), November 1980.

Cofer, C.N. Reasoning as an associative process: III The role of verbal responses in problem-solving. Journal of Genetic Psychology, 1957, Vol. 57, 55-68.

Corman, B.R. The effect of varying amounts of kinds of information as guidance in problem-solving. Psychological Monographs, 1957, Vol. 71, (2), Whole No. 431.

Cronbach, L.J., and Meehl, P.E. Construct validity in psychological tests. Psychological Bulletin, 1955, 52, 281-302.

Crutchfield, R.S., and Covington, M.V. Productive thinking tests. Berkeley, CA: Brazelton Printing Co., 1966.

Cunningham, J.D. Rigidity in children's problem-solving. School Science and Math, 1966, Vol. 66, 377-389.

Davis, G.A. Current status of research and theory in human problem-solving. Psychological Bulletin, 1966, 66, 36-58.

Dewey, J. How we think. Boston, MA: Heath, 1910.

Duncan, C.P. Recent research on human problem-solving. Psychological Bulletin, 1959, 56, 397-429.

Feldhusen, J.F., Houtz, J.C., Ringenback, S., and Lash, A. Review of problem-solving measurement tasks. Paper presented at the annual meeting of the American Psychological Association, Washington, D.C., September 1971.

Feldhusen, J.F., Houtz, J.C., and Ringenbach, S.E. The Purdue elementary problem-solving inventory. Psychological Reports, 1972, 31, 891-901.

Gagne, R.M. Problem solving. In Melton, A.W. (Ed.) Categories of Human Learning. New York, NY: Academic Press, 1964.

Gagne, R.M. The conditions of learning, (3rd ed.) New York, NY: Holt, Rinehart and Winston, 1977.

Glaser, R., Damrin, D.E., and Gardener, F.M. The tab item: A technique for the measurement of proficiency in diagnostic problem-solving tasks. Educational and Psychological Measurement, 1954, 14, 283-293.

Green, B.F. Current trends in problem solving. In Kleinmunth, B. (Ed.), Problem Solving: Research, Method and Theory. New York, NY: John Wiley and Sons, Inc., 1966.

Guilford, J.P. The nature of human intelligence. New York, NY: MaGraw-Hill, 1967.

Guilford, J.P., and Hoepfner, R. The analysis of intelligence. New York, NY: McGraw-Hill, 1971.

John, E.R. Contributions to the study of the problem solving process. Psychological Monographs, 1957, 71, (13, Whole No. 447).

Johnson, D.M. Serial analysis of verbal analogy problems. Journal of Educational Psychology, 1962, 53, 86-88.

Johnson, D.M. Solution of anagrams. Psychological Bulletin, 1966, 66, 371-384.

Katona, G. Organizing and memorizing: Studies in the psychology of learning and teaching. New York, NY: Columbia University Press, 1940.

Keisler, E.R. Teaching children to solve problems: A research goal. Journal of Research and Development in Education, 1969, 3, 3-14.

Kendler, H.H., and Kendler, T.S. Vertical and horizontal processes in problem solving. Psychological Review, 1962, 69, 1-16.

Luchins, A.S. Mechanization in problem solving. The effect of einstellung. Psychological Monographs, 1942, 54, No. 248.

Maier, N.R.F. Reasoning in humans: III. The mechanisms of equivalent stimuli and reasoning. Journal of Experimental Psychology, 1945, 35, 349-360.

McGuire, C.H., Solomon, L.M., and Forman, P.M. Clinical Simulations. New York, NY: Appleton-Century-Crofts, 1976.

Miles, D.T. Development of a test for an experimental research program in creative problem solving. Final Report Southern Illinois University, 1968.

Miller, G.A., Galanter, E., and Pribram, K.H. Plans and the structure of behavior. New York, NY: Henry Holt and Col., 1960.

Newell, A., Shaw, J.C., and Simon, H.A. The Processes of Creative Thinking in H.E. Braker and M. Westheimer, (Eds.) Contemporary Approaches to Creative Thinking. New York, NY: Atherton Press, 1962.

Newell, A., Shaw, J.C., and Simon, H.A. Elements of a theory of human problem solving. Psychological Review, 1958, 65, 151-166.

Newell, A., Simon, H.A. Human problem solving. Englewood Cliffs, NJ: Prentice-Hall, 1972.

Pribram, K.H. Languages of the Brain. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1971.

Ray, W.S. Complex tasks for use in human problem solving research. Psycholgical Bulletin, 1955, 52, 134-149.

Rimoldi, H.J.A. A technique for the study of problem-solving. Educational and Psychological Measurement, 1955, 15, 450-461.

Rimoldi, H.J.A. Problem solving as a process. Educational and Psychological Measurement, 1960, 20, 449-460.

Scandura, J.M. Problem solving, a structural/process approach. New York, NY: Academic Press, 1977.

Shann, M.H. Measuring problem solving skills and processes in elementary school children. Boston, MA: Boston University, 1976.

Simon, H.A., and Newell, A. Human problem solving: The state of the theory in 1970. American Psychologist, 1971, 26, 145-159.

Simon, R.I. Encoding effects on complex problem solving. Journal of Experimental Psychology, 1970, 83, 227-231.

Spandel, V., and Stiggins, R.J. Direct measures of writing skill: Issues and applications. Portland, OR: Northwest Regional Educational Laboratory, 1980.

Speedie, S.M., Treffinger, D.J., and Feldhusen, J.F. Teaching problem solving skills: Development of an instructional model based on human abilities related to efficient problem solving. Report to the U.S. Department of Health, Education and Welfare, Office of Education, Bureau of Research. Purdue University, August, 1973.

Stiggins, R.J. An analysis of the dimension of job-related reading. CAPT synthesis paper. Portland, OR: Clearinghouse for Applied Performance Testing, Northwest Regional Educational Laboratory, 1980.

Stiggins, R.J. Strategies for documenting and optimizing the quality of oral and practical examinations in medical education. CAPT synthesis paper. Portland, OR: Clearinghouse for Applied Performance Testing, Northwest Regional Educational Laboratory, 1979.

3220D

Tate, M.U., Stainer, B., and Harootunian, B. Differences between good and poor problem-solvers. Philadelphia, PA: School of Education, University of Pennsylvania, 1959.

Torrance, E.P. Torrance tests of creative thinking. Princeton, NJ: Personnel Press, 1966.

Treffinger, D.J. Solving problems #1 and #2. Mimeographed test used with fifth and sixth grade students. Purdue University, 1970.

Tyler, R.B. Expectancy for eventual success as a factor in problem solving behavior. Journal of Educational Psychology, 1958, 49, 166-172.

Wallas, G. The art of thought. New York, NY: Harcourt, Brace, and World, 1926.

APPENDIX

SUMMARY OF CHARACTERISTICS OF REVIEWED TESTS

PROBLEM-SOLVING TEST PROFILES

- Adult Performance Level Survey
- Analysis of Relationships
- Circus Test
- Educational Goal Attainment Test and Assessment Planning
- Learning Ability Profile
- Means-Ends Problem Solving Procedures
- Oregon Academic Ranking Test
- Purdue Problem Solving Inventory
- Ross Test of Higher Cognitive Processes
- SRA Coping Skills
- STS Educational Development Series
- Stories About Real Life Problems
- Watson-Glaser Critical Thinking Appraisal

## SUMMARY OF CHARACTERISTICS OF REVIEWED TESTS

| Test Name | Construct Orientation: Process | Construct Orientation: Skill | Problem-Solving Definition: Explicit | Problem-Solving Definition: Implicit | Item Types: Aptitude | Item Types: Achievement | Item Types: Life Skills | Response Mode: Multiple Choice | Response Mode: Original Response |
|---|---|---|---|---|---|---|---|---|---|
| Adult Perf. Level Survey | | X | | X | | | X | X | |
| Analysis of Relationships | X | | X | | | X | | X | |
| Circus Test | | X | X | | X | | | X | |
| Educ. Goal Attainment Test & Asmt. Planning | X | | | X | | | X | X | X |
| Learning Ability Profile | X | | X | | X | | | X | |
| Means-Ends Problem Solving | | X | X | | | | X | | X |
| Oregon Academic Ranking Test | X | | X | | | X | | | X |
| Purdue Problem Solving Inventory | | X | X | | | | X | X | |
| Ross Test of Higher Cognitive Processes | X | | X | | X | X | | X | |
| SRA Coping Skills | X | | | X | | | X | X | |
| STS Educational Development Series | X | | | X | | | X | X | |
| Stories About Real Life Problems | X | | | X | | | X | X | |
| Watson-Glaser Critical Thinking | | X | X | | | | X | X | |

PROBLEM-SOLVING TEST PROFILE

Name of Test: Adult Performance Level Survey  Age/Grade Level: Adult

Publisher: American College Testing  Publication Date: 1976

Conceptual definition of problem solving:
The APL Survey is not solely a test of problem-solving skill. Therefore, there is no explicit definition of problem solving and any implicit inference about examinees' problem-solving abilities would be merely conjecture. The test measures a series of skills and content areas described in the next section and scores are reported for each of examinee abilities.

Specific measurement strategy:
The APL Survey measures five skills (including identification of facts and terms, reading, writing, computation, and problem solving) in five content areas (including community resources, occupational knowledge, consumer economics, health, and government and law). The test items are multiple choice and assess various life skills. Examinees' performance is evidence of coping skills.

Role of performance assessment:
Since these items are strictly multiple choice, there is a limit to the performance component in this test. However, many of the test items measure life skill areas important to everyday living. It may be assumed that examinees who show high degrees of life skill competence may also be adept at problem solving.

Evidence of reliability:
Internal consistency estimates (KR-20 and split-half) are provided for each of the content areas and skills as well as the total survey. They range from the low 50's to the mid 60's for the subtests. Reliability estimates for the total survey equal .87.

Suggested uses of the tests:
The APL goals and objectives can be used for curriculum planning especially for adult and continuing education programs.

Evidence of validity for intended uses:
The intercorrelations among subscales in the survey range from .77 to .84.

PROBLEM-SOLVING TEST PROFILE

Name of Test: Analysis of Relationships Age/Grade Level: High School Seniors, Undergraduates and Adults

Publisher: Consulting Psychologist Press, Inc. Publication Date: 1960

Conceptual definition of problem solving:
This test is described as an intelligence test which overemphasizes verbal facility and problem solving at higher levels of ability.

Specific measurement strategy:
The analysis of relationships test is comprised of forty items, all multiple choice with no time limits given. The items are for the most part, analogies, sequences and definitions, and are primarily academic, rather than applied.

Role of performance assessment:
These forty multiple choice items rely on mental relationships among objects or ideas. As such, there is little performance required in the response notes and little real life simulation given in these stimuli.

Evidence of reliability:
Odd-even reliabilities range from .70 to .90 depending on the group.

Suggested uses of the test:
Predictive validity with (grade point averages) range from .26 to .65. Validity coefficients (for occupational) range from .22 to .76. Concurrent validity estimates range from .35 (on vocabulary measures) to .70 with the California Mental Maturity Test (an IQ test).

PROBLEM-SOLVING TEST PROFILE

Name of Test: Circus Test

Age/Grade Level: Pre-Kindergarten to Post-First Grade

Publisher: Addison-Wesley (Copyright, Educational Testing Service

Publication Date:

Conceptual definition of problem solving:
Problem solving, as defined here, requires that the child move towards the solution through a series of hypotheses and tests using convergent production. Test descriptions provided by the publisher go on to detail five sub-parts of problem solving. These include a general knowledge component, an ability to detect incongruences, an ability to define problems by categorization and variation, and the ability to evaluate solutions and implement final decisions. Problem-solving ability then is defined in terms of the abilities to utilize each of the stated problem-solving steps (regardless of student cognitive capacity).

Specific measurement strategy:
The test items are all multiple-choice and require three types of cognitive processes. The first eight problems are categorization problems, the second group of 20 problems are establishing and detecting sequence and the final group of five items require maze skills (see page 2 of the Circus User's Guide).

Role of performance assessment:
These items require the mental processes of problem solving. They are, however, typical aptitude test items and, as such, are not related to everyday problem solving and do not require any generation of answers.

Evidence of reliability:
The KR-20 Index of Consistency was used for the think-it-through (problem-solving subtest). Those reliabilities equalled .81 for Level C and Level D.

Suggested uses of the test:
This test provides both group and individual diagnostic information that can be used for program planning, program evaluation and individual student diagnosis.

Evidence of validity for intended uses:
Predictive validity equals .83 (corrected for attenuation). Evidence of construct validity was suggested for factor analysis.

PROBLEM-SOLVING TEST PROFILE

Name of Test: Educational Goal Attainment Test and Assessment Planning

Age/Grade Level: 7-12

Publisher: Phi Delta Kappa, Inc.

Publication Date: 1975

Conceptual definition of problem solving:
No specific definition for problem solving is given. However, there is an implicit understanding in these materials that problem solving is a combination of using basic intellectual aptitudes and real life experiences to solve problems encountered in everyday living.

Specific measurement strategy:
The attainment test is comprised of ten subtests dealing with various educational goals. Three of those goals have problem-solving abilities in the area of life skills, reasoning and interrelations. In all of the Education Goal Attainment tests there are multiple choice or Likert scales.

Role of performance assessment:
These items have various problem-solving tasks.

Evidence of reliability:
No evidence of reliability is provided.

Suggested uses of the test:
This battery is to be used in a multiple matrix sampling format to provide a basic understanding of the needs of student groups. This is essentially a collection of subtests that can be used in aggregate to develop a needs assessment data file for use in subsequent educational program planning.

Evidence of validity for intended uses:
No evidence of validity for any use is provided.

PROBLEM-SOLVING TEST PROFILE

Name of Test: The Learning Ability Profile  Age/Grade Level: 8 and up

Publisher: Falcon R&D Whittaker  Publication Date: 1975

Conceptual definition of problem solving:
For this test instrument, problem solving is a subset of G or learning ability. Problem solving relates to how one learns to learn, in that problems that are encountered are placed against the background of prior experience and learning, and that even most elementary inductive generalizations are context-dependent. There is a separate index of problem solving under that conceptual definition.

Specific measurement strategy:
There are 80 multiple choice items, mostly of the aptitude test type where analogy, pattern and sequence are provided and must be continued or detected by the examinee.

Role of performance assessment:
There is little in the way of a performance component to this test or the problem-solving index. Most of the processes tap spatial or verbal abilities.

Evidence of reliability:
Split-half reliabilities provide internal consistency estimates of .90 to .95.

Suggested uses of the test:
Selection and placement decisions.

Evidence of validity for intended uses:
Validity evidence is not provided.

PROBLEM-SOLVING TEST PROFILE

Name of Test: Means-Ends Problem-Solving Procedure (MEP): A Measure of Interpersonal, Cognitive Problem-solving Skill

Age/Grade Level: Adult or adolescent (requires tenth or twelfth grade level)

Publisher: Hahnemann Medical College and Hospital

Publication Date: 1977

Conceptual definition of problem solving:
For this test, problem solving is comprised of the abilities to generate alternate solutions to problems and to conceptualize means and potential obstacles in moving toward a goal (means and thinking). This "means and thinking" orientation to problem solving goes beyond cognitive problem solving for solving purely intellectual tasks; rather, "means and thinking" incorporates the notion that life problems require an empathic as well as an intellectual understanding of the problem.

Specific measurement strategy:
Ten vignettes are provided for examinees. In these vignettes, the beginning of the problem and the solution are provided; the examinee is required to fill in the middle detail that shows how the protagonist in the vignette achieves his or her own end. Examinee responses are transcribed verbatim by the examiner directly in the test booklet.

Role of performance assessment:
These ten vignettes require the examinee to solve the problem on an intellectual plane as well as develop an emotional response to the person in the vignette. Since the examinees are scored on the number of correct steps between the beginning and end of the story, an organized thinker could probably do very well. However, there is little use for actual performance other than simulated role playing.

Evidence of reliability:
Reliability estimates range from .43 to .46, depending on the sample and the administration method.

Suggested uses of the test:
This test is designed for adults and adolescents encountering problems in solving everyday problems.

Evidence of validity for intended uses:
One estimate for a group of 45 heroin addicts in a residential treatment setting obtained a correlation between staff and peer ratings of .45.

PROBLEM-SOLVING TEST REVIEW

Name of Test: Oregon Academic Ranking Test  Age/Grade Level: 3-7

Publisher: Western Psychological Services  Publication Date: 1965

Conceptual definition of problem solving:
Problem solving is operationally defined as academic or learning ability. Here the emphasis is on going beyond divergent production as a measure of "brightness" by including two other factors, creativity and abstract thinking.

Specific measurement strategy:
The OART uses eight types of test items including: making sentences, making comparisons, numbers, secret words, working problems, reasoning, completing sentences and sayings. These exercises require students to produce written or numerical solutions to problems that have many factors involved in scoring.

Role of performance assessment:
Although this test is to be used in concert with traditional IQ tests, the performance component exercises as seen in the production or generation of novel answers to requirements. The scoring rubrics are consistent with performance assessment approaches toward standardization of scoring.

Evidence of reliability:
Split-half reliability coefficient (corrected by Spearman-Brown Prophecy Formula) equals .95.

Suggested uses of the test:
The OART is an individual and group measure of academic brightness which differentiates the exceptionally bright child from the bright or average child. The OART can be used for placement into gifted programs and for research in the areas of creativity, abstract thinking, etc.

Evidence of validity for intended uses:
A concurrent study of the OART with the Stanford-Binet IQ resulted (total) in correlation of .41 at grades 3 and 7 and .82 at grade 5.

PROBLEM-SOLVING TEST PROFILE

Name of Test: Purdue Problem Solving Inventory Age/Grade Level: 2-6

Publisher: Creativity Foundation Publication Date: 1972

Conceptual definition of problem solving:
The test developers have reviewed the literature and arrived at 12 sub-tasks of problem solving, ranging from sensing that the problem exists to selecting the most unusual solution among several possible solutions. Whereas other conceptions are more process-oriented, this one is more procedural or behavioral in nature.

Specific measurement strategy:
Students are provided a slide tape set of stimuli on 49 problems. There are two to six items per subtest. In each of these vignettes the examinees are required to select from two or three distractors using their judgment of how to solve a problem (once it has been identified). This is generally a multiple-choice test.

Role of performance assessment:
This test indeed has a performance component in the sense that the stimuli are provided in a simulated setting. By using the audiovisual materials, the stimuli are more real life oriented. The response mode, however, does not require generation of problem-solving alternatives, rather examinees select among a finite list (usually three) options.

Evidence of reliability:
No reliability estimates are provided.

Suggested uses of the test:
This test is to be used for culturally disadvantaged elementary school pupils, to understand their problem-solving abilities.

Evidence of validity for intended uses:
None given.

PROBLEM-SOLVING TEST PROFILE

Name of Test: Ross Test of Higher Cognitive Processes  Age/Grade Level: 4-6

Publisher: Academic Therapy Publications  Publication Date: 1976

Conceptual definition of problem solving:
This is not so much a test of problem solving as much as it is a test of higher level thinking skills, specifically, those skills that Bloom calls analysis, synthesis and evaluation. The ability to successfully perform higher level cognitive processes is used to identify gifted students.

Specific measurement strategy:
There are eight sections, with eight to 18 items per subtest. Subtests include: analogies, deductive reasoning, missing premises, abstract relations, sequential synthesis, questioning strategies, analysis of relevant and irrelevent information, and analysis of attributes. This test has two administration periods, both lasting about an hour.

Role of performance assessment:
The items included on the Ross Test of Higher Cognitive Processes are a blend of aptitude and achievement test items. Most require piecing together bits of information to solve a problem. Performance assessment plays a part in this test because of the simulated processes required for success on this test as well as coordination of multiple sources of information to correctly solve a problem.

Evidence of reliability:
The split-half reliability coefficient (adjusted by the Spearman Brown prophecy formula) equals .92. Test re-test reliability equals .94.

Suggested uses of the test:
The Ross test may be used for the following purposes: screening students for gifted programs, evaluating gifted programs' effectiveness, and diagnosing individual student's higher level abilities.

Evidence of validity for intended uses:
The Ross test correlates with chronological age and r = .67. Construct validity was also verified in a differential performance in gifted and non-gifted students. Concurrent validity with the Lorge Thorndike Intelligence Test equals .40.

PROBLEM-SOLVING TEST PROFILE

Name of Test: SRA Coping Skills

Age/Grade Level: Junior High thru Adult

Publisher: SRA (Science Research Associates, Inc.)

Publication Date: 1978

Conceptual definition of problem solving:
The definition of problem solving provided in the SRA Coping Skills materials suggests that the examinee who does well on most of the subtests can make rational decisions based on information. This might be an operationalized definition of problem-solving abilities.

Specific measurement strategy:
The SRA Coping Skills measures eight areas. These include working, community resources, consumer economics, household management, health and safety, personal law, government and stress. Each of these subtests is administered separately or in total; then scoring and interpretation can be self-directed. The scoring ranges from "in control" to "hanging in" to "getting by" to "help!"

Role of performance assessment:
The items on the coping skills battery are very much life skills oriented and to the extent possible they present simulated problems from everyday living. The test items are multiple choice and, therefore, have little performance required in the response mode.

Evidence of reliability:
None given in this test specimen.

Suggested uses of the test:
"The coping skills program is designed to assess how well one can apply knowledge to everyday situations."

Evidence of validity for intended uses:
None given in this test specimen.

PROBLEM-SOLVING TEST REVIEW

Name of Test: STS Educational Development Series  Age/Grade Level: 4-12

Publisher: Scholastic Testing Service, Inc.  Publication Date: 1972

Conceptual definition of problem solving:
The STS Educational Development Series measures a number of facets of students' interests and abilities, one of which is solving everyday problems. This problem-solving component of the test includes problems that deal with age appropriate experiences. Operationally, the problem-solving test seems to require good judgment in selecting the best alternative from a series of distracters presented in a multiple choice format.

Specific measurement strategy:
This battery which is available at three levels (Forms A, B, and C) includes an initial assessment of student interests including career plans, school plans, and school interests, as well as two other major sections which assess nonverbal and verbal abilities. The entire test is multiple choice and subscores on 11 subtests are provided for interpretation.

Role of performance assessment:
Very little in the way of performance is required in that items are Likert scales or multiple choice items. However, many of the non-verbal and interest items measure students' awareness and interests in real life or school related areas. These often require problem solving or performance that simulate real life problem-solving ability.

Evidence of reliability:
The reliability estimates of .90 and above are provided for total test scores with subtests reliabilities in the high 80's.

Suggested uses of the test:
This is an omnibus test to be used in surveying individual and group abilities, achievements and interests.

Evidence of validity for intended uses:
"Validity studies show strong relationships between EDS and various external criterion measures." No specific validity information was provided.

PROBLEM-SOLVING TEST PROFILE

Name of Test: Stories About Real Life Problems

Age/Grade Level: Not specified (elementary grade levels)

Publisher: Northern Illinois University (pursuant to a USOE contract)

Publication Date: Not specified

Conceptual definition of problem solving:
No definition of problem solving is provided in the test specimen. However, this test suggests a definition for problem solving that would require students to understand two points of view and attempt to reconcile the two opposing positions with personal judgment based on values clarification.

Specific measurement strategy:
"Stories About Real Life Problems" consists of ten vignettes of ecological/environmental dilemmas and asks students to agree or disagree with four alternatives designed to solve the dilemma.

Role of performance assessment:
No actual performance is required of examinees. Real life problems do present problems that require students to form opinions on either side of an argument. They are not required, however, to generate alternative resolutions of the problems.

Evidence of reliability:
No reliability estimates are provided.

Suggested uses of the test:
No uses are suggested, but personal correspondence with the developer suggests that the test is to be used for understanding students' thoughts about making decisions about the environment.

Evidence of validity for intended uses:
No validity evidence is provided.

PROBLEM-SOLVING TEST PROFILE

Name of Test: Watson-Glaser Critical Thinking Appraisal

Age/Grade Level: High School and Adult

Publisher: Psychological Corporation

Publication Date: 1964

Conceptual definition of problem solving:
The critical thinking appraisal includes four tests of processes required for critical thinking. These include: inference, recognition of assumptions, deduction, interpretation and evaluation of arguments.

Specific measurement strategy:
A series of test exercises taps these four processes in terms of problems, statements, arguments and interpretations of data often found in daily life. The examinee responds to various problem-solving stems through multiple-choice responses.

Role of performance assessment:
These logic problems require examinees to simulate thought processes in the organized response patterns to problems on this test. As such, no performance is required, but examinees do perform the mental operations involved in critical thinking.

Evidence of reliability:
No evidence of reliability is provided.

Suggested uses of the test:
The critical thinking appraisal is used as a measure of critical thinking achievment for predictions in various occupational or instructional programs, for diagnosis in clinical situations, and as research for evidence of validity.

Evidence of validity for intended uses:
Concurrent validity ranges from .60 to .66 with measures of reading ability.